CERN-DD 75-4
C1

CERN-DATA HANDLING DIVISION
DD/75-4
J.M.F. Chamayou
25 March 1975

ON A DIRECT ALGORITHM FOR THE GENERATION OF LOG-NORMAL PSEUDO-RANDOM NUMBERS

(Submitted to Computing)

1. INTRODUCTION

1.1 The log-normal distribution and the energy loss distribution in atomic cascades

As a classical example of a process deeply connected with the log-normal distribution, let us give the simplest description of the losses of energy in atomic cascades [6] to illustrate the kind of distribution one must have to simulate using a random number generator.

A primary atom with an initial energy x suffers random shocks, in which it gives a part of its energy to another atom.

The density of probability of exchange of energy is, for instance, related to the hard spheres potential, i.e. uniform.

$$K_1(x,y) = \frac{1}{x} \qquad 0 \le y \le x$$
$$= 0 \qquad \text{elsewhere .} \tag{1}$$

This represents the density of probability that an atom with an initial energy x has the energy y after one collision.

Then by induction after n collisions the probability for the primary to have energy y is:

$$K_2(x,y) = \int_y^x K_1(x,z)K_1(z,y)\,dz = \frac{\log\,(x/y)}{x}$$

$$K_n(x,y) = \int_y^x K_{n-1}(x,z)K_1(z,y)\,dy =$$

$$\frac{1}{\Gamma(n)x}\left[-\log\left(\frac{y}{x}\right)\right]^{n-1}, \qquad 0 \le y \le x \tag{2}$$

$$= 0 \qquad \text{elsewhere .}$$

This process is only asymptotically log-normal, then the analogous random number generator presented here will be as convenient as an exact log-normal generator to simulate the density of probability (2).

1.2 Theoretical background

The random variable $\log X = -\log \Pi_{j=1}^{n} X_j$, where the X_j are independent and uniformly distributed on [0,1], has the following second characteristic function:

$$\psi_n(t) = -n \log (1 + it) . \tag{3}$$

Then the mean and the variance are

$$\mu = -n \quad \text{and} \quad \sigma^2 = +n \ .$$

Therefore X is asymptotically $(n \to \infty)$ log-normal $\Lambda(\mu,\sigma^2)$, i.e. it has the density

$$\frac{1}{\sqrt{2\pi}\,\sigma x} \exp\left\{-\frac{1}{2}\frac{(\log x - \mu)^2}{\sigma^2}\right\} . \tag{4}$$

But if X is $\Lambda(\mu,\sigma^2)$ and if e^a and b are constants, the random variable $Y = e^a X^b$ is $\Lambda(a + b\mu,\ b^2\sigma^2)$ (c.f. [1]). Thus, we get a method for generating pseudo-random numbers distributed log-normally $\Lambda(0,1)$.

Let us assume that

$$b = \pm \frac{1}{\sqrt{n}} \quad \text{and} \quad a = \pm \sqrt{n} \ .$$

Then, the random variable

$$\lim_{n\to\infty} Y_n = \lim_{n\to\infty} e^a \left(\prod_{j=1}^{n} X_j\right)^b \quad \text{is} \quad \Lambda(0,1) \ . \tag{5}$$

2. DISCUSSION OF THE THEORETICAL METHOD

The choice of the value of n is given by the practice; n must generally be greater than 10 for an accurate use of the central limit theorem. For instance, it is well known that an N(0,1) random variable is simulated using $\sum_{j=1}^{n} X_j - (n/2)$, where n = 12 [4].

In [2] the random variable $\sum_{i=1}^{\infty}\Pi_{j=1}^{i} X_j$ was simulated by $\sum_{i=1}^{n}\Pi_{j=1}^{i} X_j$, where n = 10.

Let us define the following random variables:

$$Z_{\pm} = e^{\pm\sqrt{n}} \left(\prod_{j=1}^{n+1} X_j\right)^{\pm 1/\sqrt{n}} . \tag{6}$$

We get the following lemma showing the accuracy of the representation of a $\Lambda(0,1)$ variable.

Lemma 1. The random variables $Z_{\pm}$ are asymptotically $(n \to \infty)$ log-normal standard variables $\Lambda(0,1)$; the order of convergence is $O(1/\sqrt{n})$.

Proof. It is well known that if {[5] see Introduction putting x = 1 in formula (2)}

$$\Xi = \prod_{j=1}^{n+1} X_j \ .$$

The density of probability of Ξ is

$$f(\xi) = \frac{(-\log \xi)^n}{n!} \qquad 0 \le \xi \le 1 \tag{7}$$

$$= 0 \qquad \text{elsewhere .}$$

Let us define

$$Z_+ = e^{\sqrt{n}}\ \Xi^{1\sqrt{n}} .$$

Then the density of Z_+ is

$$g(z_+) = \frac{(n - \sqrt{n}\ \log z_+)^n}{n!} \sqrt{n}\ e^{-n}\ z_+^{\sqrt{n-1}} \qquad 0 \le z_+ \le e^{\sqrt{n}} \tag{8}$$

$$= 0 \qquad \text{elsewhere}$$

if $n \to \infty$ by Stirling's formula [5]:

$$n! \simeq n^n\ e^{-n}\ \sqrt{2\pi n}\ (1 + \frac{1}{12n} + \ldots)$$

$$g(z_+) \simeq \frac{1}{\sqrt{2\pi}\ z_+} \left[\left(1 - \frac{1}{\sqrt{n}} \log z_+\right) \left(e^{(1/\sqrt{n})\ \log z_+}\right) \right]^n$$

but expanding $e^{(1/\sqrt{n})\ \log z_+}$, we get

$$g(z_+) \simeq \frac{1}{\sqrt{2\pi}\ z_+} \left[1 - \frac{1}{n} \left(\frac{1}{2} \log^2 z_+ + \frac{1}{3\sqrt{n}} \log^3 z_+ + \ldots \right) \right]^n$$

and we obtain easily

$$g(z_+) \simeq \frac{1}{\sqrt{2\pi}\ z_+} \exp\left\{ - \frac{\log^2 z_+}{2} \right\} \left[1 - \frac{1}{3\sqrt{n}} \log^3 z_+ + 0\left(\frac{1}{n}\right) \right] . \tag{9}$$

In the same way

$$g(z_-) \simeq \frac{1}{\sqrt{2\pi}\ z_-} \exp\left\{ - \frac{\log^2 z_-}{2} \right\} \left[1 + \frac{1}{3\sqrt{n}} \log^3 z_- + 0\left(\frac{1}{n}\right) \right] . \tag{10}$$

Q.E.D

In order to give a better approximation to the log-normal distribution, one can take advantage of the following properties

$g(z_+)$ is asymptotically > the $\Lambda(0,1)$ density for $z_+ < 1$

$g(z_+)$ is asymptotically < the $\Lambda(0,1)$ density for $z_+ > 1$

$g(z_-)$ is asymptotically < the $\Lambda(0,1)$ density for $z_- < 1$

$g(z_-)$ is asymptotically > the $\Lambda(0,1)$ density for $z_- > 1$

We therefore define the following random variable:

$$Y = Z_+ \cdot Z_- = \left(\prod_{j=1}^{n} \frac{X_i}{X_{i+n}} \right)^{1/\sqrt{n}} . \tag{11}$$

This is asymptotically a log-normal variable $\Lambda(0,2)$. Let us define

$$Z = \left(\prod_{j=1}^{n} \frac{X_i}{X_{i+n}} \right)^{b} \tag{12}$$

where $b = 1/\sqrt{2n}$, in order to get asymptotically a $\Lambda(0,1)$ variable.

Lemma 2. The random variable Z is asymptotically ($n \to \infty$) log-normal standard variable $\Lambda(0,1)$; the order of convergence is $O(1/n)$.

Proof. Using formula (3) it can be easily shown that the second characteristic function of the variable: $Y = \log Z$ is

$$\Psi_n(t) = - n \log \left(1 + \frac{t^2}{2n} \right) . \tag{13}$$

This will give rise to an $N(0,1)$ variable when $n \to \infty$; inverting $\psi_n(t)$ we get the density of Y [3].

$$g(y) = \frac{1}{\pi} \int_0^{\infty} \frac{\cos (t|y|)\, dt}{\left(1 + (t^2/2n)\right)^n} = \frac{(\sqrt{n/2}|y|)^{n-\frac{1}{2}} K_{n-\frac{1}{2}} (\sqrt{2n}|y|)}{\Gamma(n)\sqrt{\pi}} =$$

$$\frac{\sqrt{2n}\, e^{-\sqrt{2n}|y|}}{2^{2n-1}\, \Gamma(n)} \sum_{k=0}^{n-1} \frac{\Gamma(2n - k - 1)(2\sqrt{2n}|y|)^k}{\Gamma(n - k)\Gamma(k + 1)} \qquad -\infty < y < \infty . \tag{14}$$

Where $K_r(x)$ is the Bessel function of second order and imaginary argument [8]. Then the density of Z is

$$h(z) = \frac{1}{\pi z} \int_0^{\infty} \frac{\cos (t|\log z|)\, dt}{\left(1 + (t^2/2n)\right)^n} \qquad 0 \leq z < \infty$$

$$= 0 \qquad\qquad \text{elsewhere}$$

and

$$h(z) \approx \frac{1}{\pi z} \int_0^\infty e^{-t^2/2} \cos (t|\log z|)\, dt \cdot \left[1 + 0\left(\frac{1}{n}\right)\right]$$

$$= \frac{1}{\sqrt{2\pi}\, z} \exp \left\{- \frac{\log^2 z}{2}\right\} \left[1 + 0\left(\frac{1}{n}\right)\right] \qquad (15)$$

Q.E.D.

3. THE ALGORITHM

In FORTRAN IV the algorithm corresponding to the random variable (12) is the following:

```
      N=***
      AN=1.
      BN=1./SQRT(2.*FLOAT(N))
      -----------------------
      Z=RLONOR(N,AN,BN)
      -----------------------
      FUNCTION RLONOR(N,AN,BN)
      X=1.
      DO 1 I=1,N
    1 X=X*RNDM(I)/RNDM(I+N)
      RLONOR=AN*X**(+BN)
      RETURN
      END
```

where RNDM(I) is a function generating pseudo-random uniform numbers (a congruential method is used for the tests). AN and BN can be adjusted in order to get a $\Lambda(\mu,\sigma^2)$ variable

```
      AN=EXP (μ)
      BN=(...)*σ .
```

4. STATISTICAL TESTS

The mean of a log-normal distribution $\Lambda(0,1)$ is [1]

$$\alpha = e^{\frac{1}{2}} = 1.6487 .$$

But

$$\bar{\alpha} = \underbrace{\int_0^1 \cdots \int_0^1}_{2n} \left(\frac{x_1 x_2 \cdots x_n}{x_{n+1} \cdots x_{2n}} \right)^{1/\sqrt{2n}} = dx_1 \ldots dx_{2n} = \left(1 - \frac{1}{2n}\right)^{-n} = \alpha \left[1 + 0\left(\frac{1}{n}\right)\right], \tag{16}$$

see Table 1.

Figure 1 shows the comparison between the density of a $\Lambda(0,1)$ variable and the histogram of the pseudo-random numbers; it is unimodally shaped with a modal value corresponding to e^{-1} [1].

On a CDC 7600 computer, the time of calculation is

T_1 = (14 + 5n) μsec/log-normal pseudo-random number

T_2 = (10 + 5n) μsec/normal pseudo-random number.

This generator has been tested using some one sample non-parametric tests: Kolmogorov's test, Cramer-Von Mises test and Renyi's test. The results are presented in Table 2. Comparison is made with

(*) the Box-Muller method [4],

(**) the sum of 12 random uniform deviates,

(***) 2000 true normal random numbers [7].

[It can be noted that (**) and n = 6 are tested with the same series of random uniform numbers for each pseudo-random log-normal number].

5. CONCLUSION: Interpretation of the results

The interest of this pseudo-random number generator is to provide directly $\Lambda(0,1)$ variables without the help of any $N(0,1)$ variable.

Normal deviate generators provide slightly quicker the $\Lambda(0,1)$ random variables for an identical bias and the same test confidence level, as can be seen in the table (n = 5,6). But, in fact, many of the processes involved in the log-normal distributions are described in the same terms as in the example given in the Introduction.

Then these processes are only asymptotically log-normal, and if one wants to simulate that kind of distribution an analogous random generator will be as convenient as the exact log-normal. For instance, the density of probability of the energy of a primary atom (2) that suffers n = 36 shocks will be closely represented using the random variable (12) where n = 6.

Table 1

n	$\mu = 0$	$\sigma^2 = 1$	$K_3 = 0$	$\alpha = 1.649$	T µsec
4	-0.002	0.99	0.008	1.716	34
5	0.003	1.008	-0.008	1.696	39
6	0.007	1.018	0.018	1.718	44
7	-0.005	0.991	-0.019	1.657	49
8	-0.003	1.007	0.003	1.673	54
(*)	-0.003	0.989	0.013	1.637	39
(**)	-0.005	1.002	0.014	1.646	35
(***)	-0.018	0.992	0.052	1.620	-

m = 20000 pseudo-random numbers

Table 2

Kolmogorov's test				Cramer's test	
n	D_m	z	$Pr\left[\sqrt{m}\, D_m \geq z\right]$ %	W_m^2	$Pr\left[W_m^2 \leq w_m^2\right]$ %
4	0.026	1.14	14.5	0.30	86.8
5	0.022	1.0	26.5	0.22	76.9
6	0.031	1.4	40.2	0.133	55.3
7	0.018	0.82	52.0	0.173	67.3
8	0.016	0.72	68.4	0.1	41.3
(*)	0.024	1.06	21.1	0.138	57.1
(**)	0.018	0.83	49.5	0.130	54.4
(***)	0.026	1.16	13.7	0.144	59.2

Renyi's test a = 0.1			
n	D_m a)	z	$Pr\left[D_m \sqrt{\frac{ma}{1-a}} \leq z\right]$ %
4	0.105	1.56	76.5
5	0.077	1.14	49.3
6	0.058	0.86	23.8
7	0.095	1.41	68.3
8	0.078	1.16	51.1
(*)	0.047	0.71	10.8
(**)	0.051	0.76	15.3
(***)	0.102	1.51	74.0

m = 2000 pseudo-random numbers

a) $D_m = \max\limits_{0<a\leq F_m\leq 1} \left|\frac{F_m - S_m}{F_m}\right|$

REFERENCES

[1] J. Aitchison and J.A.C. Brown, The log-normal distribution (Cambridge University Press, Cambridge, 1966).

[2] J.M.F. Chamayou, Math. Comput. 27, 197 (1973).

[3] I.S. Gradshteyn and I.M. Ryzhik, Table of integrals, Academic Press, New York, 1965).

[4] B. Jansson, Random number generators (Pettersons, Stockholm, 1966).

[5] M.G. Kendall and A. Stuart, The advanced theory of statistics (C. Griffin, London, 1963), Vol. 1.

[6] G.M. Kinchin and R.S. Pease, Rep. Progr. Phys. 18, 1 (1955).

[7] A million random digits (The Free Press, The Rand Corporation, Glencoe, Ill., 1955.)

[8] K. Pearson, S.A. Stouffer and F.N. David, Biometrika 24, 293 (1932).

Figure caption

Fig. 1 : Log-normal distribution

* Log-normal density

▲ Density of the pseudo-random numbers (n = 6)

\- Histogram of the pseudo-random numbers (m = 9000, n = 6)

Dashed area: difference between the random histogram and the exact histogram (i.e. each bin is the value of the integral of the density in the bin).

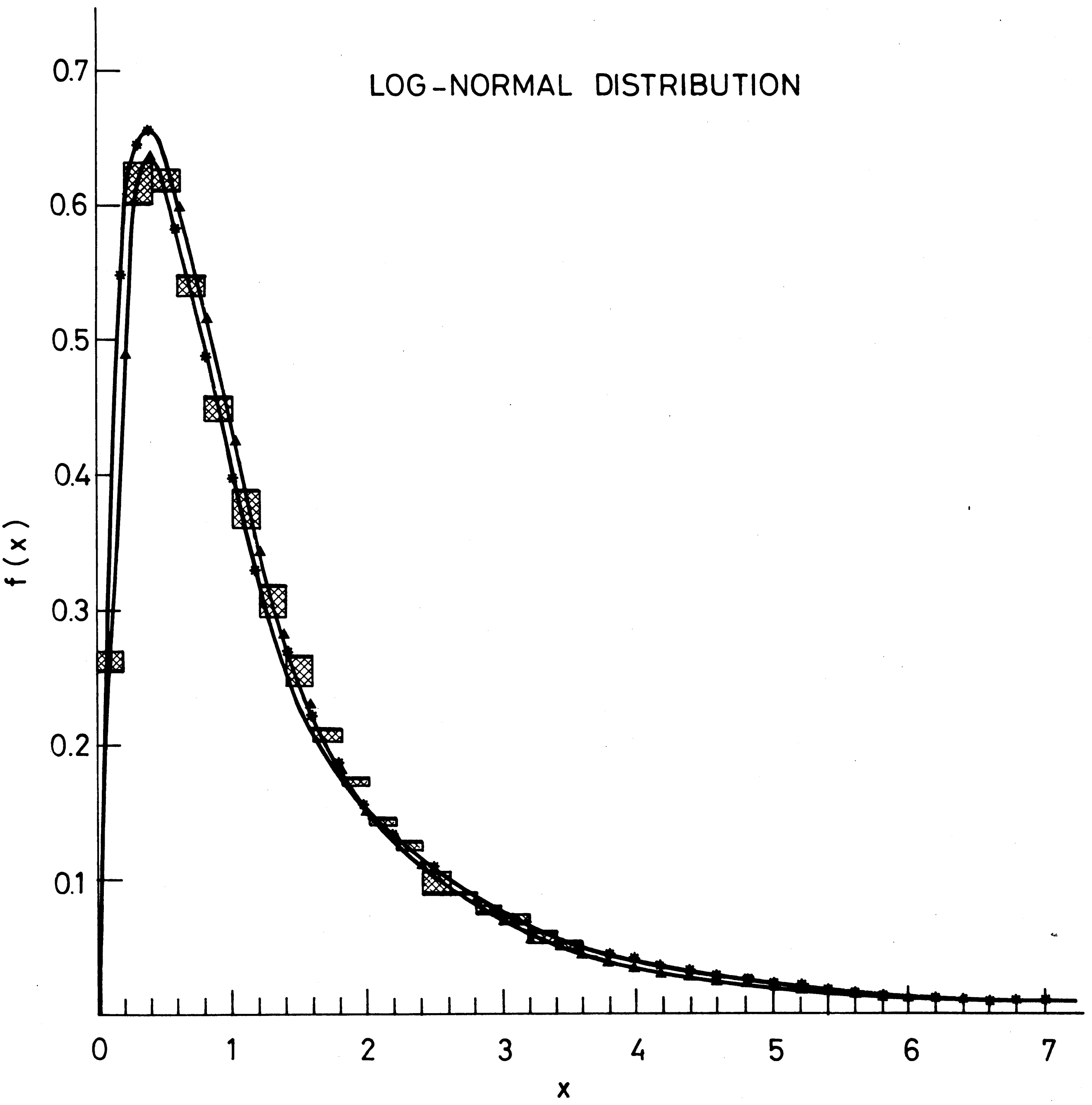

Fig. 1